

学校编码: 10384

密级_____

学号: 22420091151147

厦 门 大 学

硕 士 学 位 论 文

杂色鲍早期发育阶段的转录组学分析

Transcriptomic analysis on early developmental stages of

Haliotis diversicolor

黄子夏

指导教师姓名: 陈 军 副教授

专 业 名 称: 海 洋 生 物 学

论文提交日期: 2012 年 6 月

论文答辩时间: 2012 年 6 月

2012年 6月

厦门大学博士论文摘要库

厦门大学学位论文原创性声明

本人呈交的学位论文是本人在导师指导下,独立完成的研究成果。本人在论文写作中参考其他个人或集体已经发表的研究成果,均在文中以适当方式明确标明,并符合法律规范和《厦门大学研究生学术活动规范(试行)》。

另外,该学位论文为()课题(组)的研究成果,获得()课题(组)经费或实验室的资助,在()实验室完成。(请在以上括号内填写课题或课题组负责人或实验室名称,未有此项声明内容的,可以不作特别声明。)

声明人(签名):

年 月 日

厦门大学博硕士论文摘要库

厦门大学学位论文著作权使用声明

本人同意厦门大学根据《中华人民共和国学位条例暂行实施办法》等规定保留和使用此学位论文，并向主管部门或其指定机构送交学位论文（包括纸质版和电子版），允许学位论文进入厦门大学图书馆及其数据库被查阅、借阅。本人同意厦门大学将学位论文加入全国博士、硕士学位论文共建单位数据库进行检索，将学位论文的标题和摘要汇编出版，采用影印、缩印或者其它方式合理复制学位论文。

本学位论文属于：

（ ） 1. 经厦门大学保密委员会审查核定的保密学位论文，
于 年 月 日解密，解密后适用上述授权。

（ ） 2. 不保密，适用上述授权。

（请在以上相应括号内打“√”或填上相应内容。保密学位论文应是已经厦门大学保密委员会审定过的学位论文，未经厦门大学保密委员会审定的学位论文均为公开学位论文。此声明栏不填写的，默认为公开学位论文，均适用上述授权。）

声明人（签名）：

年 月

厦门大学博硕士论文摘要库

目录

摘要	I
ABSTRACT	III
第 1 章 绪论	1
1.1 海洋贝类生物早期发育研究概况	1
1.1.1 软体动物的基本特征	1
1.1.2 我国贝类养殖业的现状和遇到的挑战	2
1.1.3 贝类生物幼体附着变态研究现状和发展趋势	2
1.2 鲍的早期胚胎发育研究概况	5
1.2.1 鲍早期胚胎发育生物学过程	5
1.2.2 鲍的早期发育研究进展	6
1.3 DNA 测序技术的历史与发展现状	7
1.3.1 经典 DNA 测序技术	8
1.3.2 454 高通量测序技术	9
1.3.3 Solexa 高通量测序技术	10
1.3.4 ABI SOLiD 高通量测序技术	10
1.3.5 第三代测序技术	11
1.3.6 高通量测序技术的应用	13
1.4 转录组研究方法	14
1.4.1 基于杂交的转录组研究方法	15
1.4.2 基于 Sanger 测序的转录组研究方法	15
1.4.3 基于高通量测序 (RNA-Seq) 的转录组分析方法	16
1.4.4 转录组研究方法在软体动物研究中的应用	17
1.5 本研究的目的、意义和内容	18
第 2 章 杂色鲍 454 数据预处理拼接以及 Unigene 构建	19
2.1 生物信息学分析平台构建	19
2.1.1 Linux 虚拟机平台搭建	19
2.1.2 应用程序的下载与安装	20

2.1.3 生物数据库及序列的下载与获取.....	22
2.2 杂色鲍八个时期文库序列预处理、拼接、Unigene 的构建.....	23
2.2.1 杂色鲍八个文库构建与测序.....	23
2.2.2 杂色鲍各个时期序列的预处理.....	24
2.2.3 序列拼接.....	26
2.2.4 Scaffold 构建.....	27
2.2.5 不同物种转录组的比较.....	28
2.3 结果与分析.....	28
2.3.1 序列拼接方法的分析与讨论.....	28
2.3.2 杂色鲍幼体转录组序列的拼接结果与讨论.....	30
2.3.3 Scaffold 和 Singleton 的 PCR 验证.....	32
2.3.4 杂色鲍转录组与其他鲍类转录组的比较.....	34
第 3 章 杂色鲍 Unigene 的基因注释和基因本体论 (Gene Ontology)	
注释.....	36
3.1 Unigene 的基因注释方法构建.....	36
3.2 Gene Ontology (GO) 注释方法构建.....	36
3.2.1 Gene Ontology (GO) 注释背景介绍.....	36
3.2.2 Gene Ontology (GO) 注释方法流程.....	37
3.2.3 Gene Ontology (GO) 注释方法评估.....	39
3.3 结果与分析.....	40
3.3.1 杂色鲍 Unigene 基因注释结果与分析.....	40
3.3.2 GO 注释方法构建的讨论.....	42
3.3.3 杂色鲍 Unigene 的 GO 注释结果与分析.....	43
第 4 章 杂色鲍幼体基因表达谱的构建.....	46
4.1 杂色鲍幼体最高表达量基因和基因表达谱的方法构建.....	46
4.1.1 最高表达量基因和基因表达谱的方法构建流程.....	46
4.1.2 杂色鲍各个时期序列的预处理与去冗余.....	47
4.1.3 杂色鲍各个时期序列的相似性比对以及最高表达基因的获得.....	48

4.1.4 杂色鲍幼体转录组基因表达谱的构建.....	49
4.1.5 基因表达谱的实验验证方法.....	49
4.1.6 基因表达的聚类分析方法构建.....	50
4.2 结果与分析.....	50
4.2.1 杂色鲍幼体表达量最高的 20 个基因.....	50
4.2.2 基因表达谱的实时定量 PCR 实验验证.....	53
4.2.3 基因表达谱构建方法优缺点讨论.....	55
4.2.4 爆发性基因.....	56
4.2.5 30 个表达波动最剧烈的基因.....	57
4.2.6 基因表达谱的聚类分析.....	61
第 5 章 结论与展望.....	64
5.1 结论.....	64
5.2 创新点.....	64
5.3 不足之处.....	65
5.4 研究展望.....	65
参考文献.....	66
攻读硕士学位期间发表的论文.....	75
致 谢.....	76

厦门大学博硕士论文摘要库

TABLE OF CONTENTS

ABSTRACT(in CHINESE)	I
ABSTRACT(in ENGLISH)	III
CHAPTER 1 PREFACE	1
1.1 Research summary of early development of marine shellfish	1
1.1.1 Basic features of molluscan species.....	1
1.1.2 Status quo and challenge of shellfish aquaculture in China	2
1.1.3 Status quo and trend of settlement and metamorphosis research of shellfish.....	2
1.2 Research summary of early embryonic development in abalone	5
1.2.1 Biological processes of early embryonic development in <i>Haliotis diversicolor</i>	5
1.2.2 Reseach progress of early development in abalone	6
1.3 History and status quo of DNA sequencing technology	7
1.3.1 Classical DNA sequencing technology	8
1.3.2 454 high-throughput sequencing technology.....	9
1.3.3 Solexa high-throughput sequencing technology	10
1.3.4 ABI SOLiD high-throughput sequencing technology	10
1.3.5 The third generation sequencing technology	11
1.3.6 Application of high-throughput sequencing technology.....	13
1.4 Research methods of transcriptome	14
1.4.1 Transcriptomic methods based on hybridization	15
1.4.2 Transcriptomic methods based on Sanger sequencing	15
1.4.3 Transcritpomic methods based on high-throughput sequencing (RNA-Seq).....	16
1.4.4 Application of transcriptomic methods in researches of molluscan species	17
1.5 Purpose, significance and contents of this study	18

CHAPTER 2 PRETREATMENT, ASSEMBLY AND UNIGENE

CONSTRUCTION OF 454 DATA OF *HALIOTIS*

DIVERSICOLOR 19

2.1 Construction of bioinformatic analysis platform..... 19

2.1.1 Construction of Linux virtual machine platform 19

2.1.2 Download and installation of software 20

2.1.3 Download and acquisition of biological databases and sequences 22

2.2 Sequence pretreatment, assembly and Unigene construction of seven libraries of *Haliotis diversicolor* 23

2.2.1 Construction and sequencing of eight libraries of *Haliotis diversicolor* 23

2.2.2 Sequence pretreatment of different libraries of *Haliotis diversicolor* .. 24

2.2.3 Sequence assembly 26

2.2.4 Scaffold construction 27

2.2.5 Transcriptome comparison among different species..... 28

2.3 Result and analysis..... 29

2.3.1 Analysis and discussion of assembly methods..... 29

2.3.2 Result and discussion of assembly of *Haliotis diversicolor* larval transcriptomes 30

2.3.3 PCR validation of Scaffold and Singleton 32

2.3.4 Transcriptome comparison between *Haliotis diversicolor* and other abalone species..... 34

CHAPTER 3 GENE ANNOTATION and Gene ONTOLOGY

ANNOTATION OF UNIGENE 36

3.1 Method construction of gene annotation of Unigene 36

3.2 Method construction of Gene Ontology (GO) annotation..... 36

3.2.1 Background introduction of Gene Ontology (GO) annotation 36

3.2.2 Method process of Gene Ontology (GO) annotation 37

3.2.3 Method assessment of Gene Ontology (GO) annotation	39
3.3 Result and analysis.....	40
3.3.1 Result and analysis of gene annotation of Unigene in <i>Haliotis diversicolor</i>	40
3.3.2 Discussion of method of GO annotation.....	42
3.3.3 Result and analysis of GO annotation of Unigene in <i>Haliotis diversicolor</i>	43
CHAPTER 4 CONSTRUCTION OF GENE EXPRESSION	
PROFILE OF <i>HALIOTIS DIVERISCOLOR</i> LARVAE	46
4.1 Top highly expressed genes and method construction of gene expression profile	46
4.1.1 Construction processes of top highly expressed genes and gene expression profile.....	46
4.1.2 Sequence pretreatment and redundancy removal of each abalone library	47
4.1.3 Sequence similarity alignment and acquisition of mostly expressed genes of each abalone library.....	48
4.1.4 Construction of transcriptomic gene expression profile of larval abalone	49
4.1.5 Experimental methods of gene expression profile.....	49
4.1.6 Method construction of cluster analysis of gene expression	50
4.2 Result and analysis.....	50
4.2.1 20 top expressed gene in abalone larvae.....	50
4.2.2 Real time PCR experimental validation of gene expression profile	53
4.2.3 Discussion of advantage and disadvantage of gene expression profile method.....	55
4.2.4 Explosive genes	56
4.2.5 30 mostly fluctuated genes	57
4.2.6 Cluster analysis of gene expression	61

CHAPTER 5 CONCLUSION AND PROSPECT	64
5.1 Conclusion	64
5.2 Innovation.....	64
5.3 Disadvantage	65
5.4 Research outlook	65
REFERENCE.....	66
PUBLISHED PAPER	75
ACKNOWLEDGEMENTS	76

摘要

海洋无脊椎动物的附着和变态过程是一种常见的海洋生态学现象,然而基因背景的缺乏严重制约了附着变态分子机理的研究。杂色鲍 (*Haliotis diversicolor*) 作为软体动物腹足类的重要物种,不仅是水产养殖重要经济贝类,还是良好的研究海洋无脊椎动物幼体变态发育模型。因此,对杂色鲍早期发育机理的阐述具有重要的理论和实践意义。但是,目前杂色鲍早期发育分子机制的研究手段十分低效且早期发育相关基因序列相对匮乏。因此本研究针对这个需求,构建了杂色鲍 8 个时期转录组文库,并利用 454 测序技术对其进行测序,获得了宏量的有关杂色鲍幼体发育的基因序列。

本研究采集了杂色鲍早期发育 7 个关键时期的样品,即二细胞期、桑葚期、担轮幼虫时期、面盘幼虫中期、面盘幼虫后期、感受态附着失败幼体和围口壳幼体期,以及肠道组织样品。经过一系列转录组文库构建流程,利用 454 GS FLX Titanium 平台对 8 个杂色鲍转录组文库进行测序,共获得 366,991 条序列。利用生物信息学方法对序列进行预处理、去冗余、拼接和 Scaffold 构建等过程,共得到 35,415 个 Unigenes,其中 701 个 Scaffold, 9,567 个 Contig 以及 25,147 个 Singleton; 通过序列相似性比对, 9,513 (26.9%) 条 Unigene 在 Swissprot 或 Nr 蛋白数据库中找到相似性序列。同时,本研究还开发集成了适用于小型实验室的大规模 GO 注释生物信息学流程,通过对杂色鲍幼体转录组序列进行分析,发现其中 7,566 条序列被分配到 GO 词条。通过与南非鲍成体和文蛤幼体 GO 注释的 Spearman 相关系数分析比较,发现杂色鲍幼体的基因组成结构与文蛤幼体更加接近,说明不同软体动物的幼体发育很可能具有相似的生物代谢过程。

本研究分析了杂色鲍幼体表达量最高的 20 个基因以及这些基因的功能。通过分析我们发现,这些基因与加州海兔幼体表达量最高的基因组成十分相似,共有 11 个基因存在交集,因此推测不同软体动物幼体发育的生物过程十分相似。同时,我们还发现了 3 未知功能的基因,这说软体动物幼体的某些基本发育过程还鲜为人知。本研究同时开发创新了基因数字表达谱的构建方法,获得了近 1,500 个基因的表达模式,从组学的角度研究探讨杂色鲍幼体变态发育的分子机理。通

过实时定量 PCR 对 20 个基因的表达水平的验证,证明基因表达谱数值准确可信,为杂色鲍附着变态分子机理的研究提供了良好的参考。通过对表达量最为波动 30 个基因的分析 and “爆发性基因”数目的统计,我们发现“爆发性基因”主要集中于二细胞期、担轮幼虫期和围口壳幼体期,说明杂色鲍幼体在这三个发育过程存在巨大的生理生态转变,而“爆发性基因”很可能与杂色鲍的变态附着密切相关。同时,通过聚类分析的方法从整体上对特定基因在不同幼体时期中的表达水平进行了评估和探讨。

本研究不仅在很大程度上弥补了杂色鲍早期发育基因资源相对匮乏的空缺,对转录组的生物信息学分析手段进行了开发和创新,而且获得了杂色鲍幼体变态附着密切相关的基因,站在更高的角度上探讨其分子机制。因此,本文对杂色鲍早期发育的转录组分析,可以成为其它海洋贝类动物基因资源积累和开发的参考。

关键词: 杂色鲍; 高通量测序; 转录组学; 生物信息学; 早期胚胎发育; 附着变态; 基因差异表达

Degree papers are in the "[Xiamen University Electronic Theses and Dissertations Database](#)". Full texts are available in the following ways:

1. If your library is a CALIS member libraries, please log on <http://etd.calis.edu.cn/> and submit requests online, or consult the interlibrary loan department in your library.
2. For users of non-CALIS member libraries, please mail to etd@xmu.edu.cn for delivery details.

厦门大学博硕士论文摘要库